DETERMINING DISPLACEMENT FIELDS ALONG CONTOURS FROM IMAGE SEQUENCES

Patrick BOUTHEMY

IRISA / INRIA
Campus de Beaulieu
35042 RENNES, France

ABSTRACT

We propose a framework for image flow analysis consisting of three major stages; i.e.:

- determine moving edges by some local process;
- integrate motion information along linked contours;
- propagate motion estimation through homogeneous regions obtained inside these contours.

This paper is concerned with the two first stages. A procedure, based on some local modeling and maximum likelihood scheme, bas been designed to perform the first step. After some linking process, constraints provided by the measurements gained from the first stage can be combined to compute the velocity field along contours, by minimizing some simple functional. To this end, a gradient algorithm is used with a recursive estimation from one point to its successor in the chain.

RESUME

Nous proposons un schéma d'obtention du champ des vitesses dans une séquence d'images s'articulant en trois étapes, à savoir :

- déterminer localement les éléments de contour en mouvement;
- intégrer l'information de mouvement le long des lignes contours chaînées;
- propager l'estimation du mouvement à l'intérieur des zones homogènes délimitées par ces lignes contours.

Le papier traite des deux premières étapes. Une procédure, basée sur une modélisation locale et un critère de maximum de vraisemblance, a été conçue afin de réaliser le premier point. Après chaînage, les mesures issues du premier niveau peuvent être combinées afin de calculer le champ des vitesses complet le long des lignes contours via la minimisation d'une fonctionnelle simple. A cette fin, un algorithme de gradient est mis en oeuvre avec une récurrence de point en point le long de la chaîne contour.

KEYWORDS: image sequence, moving edge determination, motion estimation, local modeling, maximum likelihood test, stochastic gradient.

I INTRODUCTION

Image sequence analysis has received more and more attention since 70's. In particular substantial studies have been concerned with motion estimation accross changing two-dimensional images. Two main motivations have subtented there research efforts. First, motion computation represents an attractive challenge in order to design some robust, tractable and general-purpose method. On the other hand, application areas never stop broadening, |1|.

Meteorological applications (determining wind fields owing to cloud motion estimation, |2|), military domain (target tracking) were among pioneer ones. Then came interframe image coding for broadcast television or videoconferencing purpose |3,4|. For a few years, other potential applications have appeared: biomedical (e.g., angiocardiography |5|,) robotics (mobile robot, |6|), traffic monitoring, graphics ... These new domains are not only interested in two-dimensional motion as it is, but as intrinsic features conveying information about the depicted 3D-scene. Indeed motion in the imaging plane provides primary cues to relative depth, structure and 3D-movements of objects in space |7,8|.

The motion in the imaging plane is usually referred to as the "optic flow". Optic flow can be represented as a vector field: the field of apparent velocities of brightness patterns in the image due to relative motion of camera and objects in space. (As one's uses a dicrete representation of an image sequence, displacement vector fields and velocity vector fields are usually confused, although mathematically of different nature).

Discriminating discontinuities in the velocity field is a key problem in motion estimation schemes whatever they are. Indeed, feature-based methods require cooperative matching procedures |9|, and gradient-based methods involve some smoothing constraint |10,11|. Thus, we have designed a method whose first task is to cope with these discontinuities, which are tied to contours in the image, such as occluding contours, joint ones ...

We propose a framework for image flow analysis consisting of three major stages; i.e.:
determine moving edges by some local process;
link these eges and integrate motion information along contours;
propagate motion estimation through homogeneous re-

gions obtained inside these contours.

The two first issues are addressed in this paper.

The first step can be considered as an early processing whose output contains location and spatial direction of an edge element and component of its displacement in the direction perpendicular to the local orientation of the edge. It is well-known that only such partial motion information can be reached by local operations (this point is often referred to as the aperture problem). A maximum likelihood method, based on some local modeling has been designed for this purpose.

Then constraints provided by these measurements gained from the first stage can be combined to compute the velocity field along contours, if however variations in spatial orientations occur along such contours. This computation results from the minimization of some simple functional by a stochastic gradient algorithm.

II LOCAL DETERMINATION OF A MOVING EDGE

II.1 - Modeling of a moving edge

An image sequence is considered as a 3D-space (x,y,t). A spatial 2D-edge in an image is modeled as a small local linear segment. Hence a moving 2D-edge is locally modeled as a small planar patch in the spatio-temporal 3D-space (x,y,t). The direction θ (w.r.t. the x-axis) of the 2D-edge centered in (x_0,y_0) in the xy-plane at time t_0 and its velocity $\vec{V} = (\frac{dx}{dt}, \frac{dy}{dt}, 1)$ determine the orientation of this planar patch (see figure 1). This planar modeling is equivalent to the first order approximation that most gradient-based methods take into account.

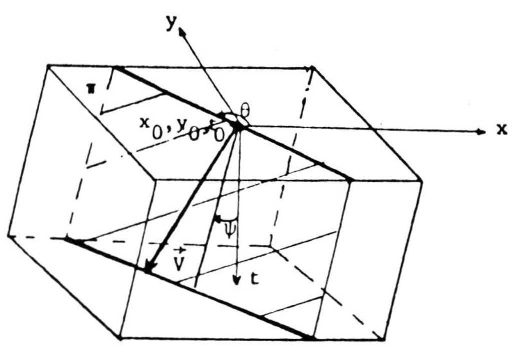

Figure 1: *Local modeling of a moving edge as a planar patch.*
$\theta \in [0,\Pi[\ ;\ \psi \in [0,\Pi/2[$

Let us consider an elementary volume Π, in the 3D-space (x,y,t), located around point $\ell=(x_0,y_0,t_0)$. Two hypotheses (or local configurations) can be acting:

H_0: there is no spatio-temporal edge inside Π; then the intensity distribution within Π is modeled as c_0+b, where c_0 is a constant and b denotes a zero-mean Gaussian noise with variance σ^2.

H_1: there is a spatio-temporal edge inside Π, i.e. a small planar patch P splitting Π between two sub-regions, Π_1 and Π_2. Then the intensity distribution is modeled as: c_1+b within Π_1; c_2+b within Π_2, where c_1 and c_2 are two different constants.

The orientation of the planar patch can be defined by the two following angles: θ (w.r.t. to the x-axis) and ψ (w.r.t. to the t-axis) as illustrated in Figure 1. The component $V^{\perp}$ of $\vec{V}$ perpendicular to the spatial 2D-edge and projected in plane $t=t_0$, is given by: $V^{\perp} = \tan\psi$. It is obvious that only this component $V^{\perp}$ can be inferred from the local determination of this planar patch. Note that the case of a static edge belongs to hypothesis H_1; $\vec{V}=(0,0,1)$ and $\psi=0$. Indeed such an edge will be considered as a "moving" edge, whose displacement is zero.

The problem now is how to select one hypothesis versus the other one. The test in order to decide between these two hypotheses will be designed using some maximum likelihood scheme.

II.2 - Maximum likelihood test

Details in mathematical developments can be found in |12|, concerning the maximum likelihood test designed for detecting moving edges along with estimating their parameters. It is expressed by:

$$\max_{\ell,\theta,\psi}\ \max_{c_1,c_2}\ \min_{c_0}\ \mathrm{LRV} \geqq \lambda \qquad (1),$$

where LRV is the log-ratio of likelihood functions L_1 and L_0, respectively associated with hypotheses H_1 and H_0. The likelihood function is merely the joint probability density function of the intensities within elementary volume Π. It is easily derived as Gaussian distributions are involved and independent intensity random variables are assumed. λ is a predetermined threshold.

Clearly, hypothesis H_1 is selected if the obtained maximum value of LRV, is greater than λ. Then one can conclude that a moving edge is located at $\hat{\ell}$ with spatial direction $\hat{\theta}$ and "perpendicular" velocity $V^{\perp} = \tan\hat{\psi}$, where $\hat{\ell}$, $\hat{\theta}$, $\hat{\psi}$ are precisely values of (ℓ,θ,ψ) which have satisfied the mentioned criterion (1).

Yet one problem arises. No analytical closed-forms can be derived to express the optimal estimators $\hat{\theta},\hat{\psi}$ corresponding to the geometrical characteristics of the model. Thus a predefined set of given configurations Φ_j, j=1,..., G, will be considered.

For a given geometric configuration (θ_j,ψ_j), the optimal estimators $\hat{c}_i$ (i=0,1,2) concerning intensity aspects satisfy:

$$\frac{\partial \mathrm{LRV}(\ell,\Phi_j)}{\partial c_i} = 0$$

which leads to

$$\hat{c}_0 = \frac{1}{n}\sum_{p\in\Pi} f(p)\ ;\ \hat{c}_1 = \frac{1}{n_1}\sum_{p\in\Pi_1} f(p)\ ;\ \hat{c}_2 = \frac{1}{n_2}\sum_{p\in\Pi_2} f(p) \qquad (2),$$

where f(p) are observed intensity values within Π, n (resp. n_1,n_2) is the number of points within Π; (resp. Π_1,Π_2).

II.3 - Computational scheme

It turns out, |12|, that maximizing LRV comes to maximizing the following expression:

$$\mathrm{CRV}\ (\ell,\Phi_j) = \sqrt{\frac{n_1 n_2}{2n}}\ |\hat{c}_1-\hat{c}_2| \tag{3}$$

Using (2) and (3), we can write function $\mathrm{CRV}(\ell,\Phi_j)$ in the form:

$$\mathrm{CRV}(\ell,\Phi_j) = \Big|\sum_{m\in M} a_j(m)\ f(\ell+m)\Big| \tag{4}$$

where M is a set of vectorial indices such that $\{\ell+m,\ m\in M\}$ represent all points of volume Π, and coefficients a_j's only depend on a predefined configuration Φ_j. Indeed, the computational implementation of this local estimation process merely consists of convolution operations.

The process for determining moving edges between two successive images can be summarized as follows. For each point ℓ in the first image:

. calculate $\mathrm{LRV}(\ell,\Phi_j)$ for each mask $\{a_j\}$ corresponding to geometric configuration Φ_j

. select orientation Φ_k which maximizes function LRV; if $\mathrm{LRV}(\ell,\Phi_k)\geq\lambda$, then a moving edge is said to be potentially present at point ℓ, whose estimated parameters are Φ_k and associated likelihood value denoted by $\mathrm{CONF}(\ell)=\mathrm{LRV}(\ell,\Phi_k)$; else no moving edge is determined and $\mathrm{CONF}(\ell)$ is set to 0.

For each previously selected point ℓ
- If $\mathrm{CONF}(\ell)>\mathrm{CONF}(\ell_1)$ and $\mathrm{CONF}(\ell)>\mathrm{CONF}(\ell_2)$, where points ℓ_1 and ℓ_2 are the two neighbours of ℓ in the direction perpendicular to θ_k, then conclude that a moving edge is located at $\hat{\ell}=\ell$, whose parameters are given by $\hat{\Phi}=\Phi_k$.

This last step could be interpreted as a thinning procedure. Indeed it corresponds to the local maximization of CRV subject to location parameter ℓ as expressed in (1).

If the volume Π intersects I images, CRV can be decomposed into:

$$\mathrm{CRV}(\ell,\Phi_j)=\Big|\sum_{i\in I}\sum_{m_i\in M_i} a_j(m_i)f(\ell+m_i)\Big|$$

$$=\Big|\sum_{i\in I}\mathrm{CRVi}\ (\ell,\Phi_j)\Big| \tag{5}$$

where M_i is a set of vectorial indices such that $\{\ell+m_i,\ m_i\in M_i\}$ represent all points belonging to volume Π and image i. Hence, this approach can embrass with the same formalism cases where two and more images are considered.

An additional heuristic is introduced to avoid false detections. Before concluding that a spatio-temporal edge is present at point ℓ according to the criterion (1), the following constraint must be satisfied:

$$\mu_1 \leq |\mathrm{CRV1}|\ /\ |\mathrm{CRVi}| \leq \mu_2,\ \text{for all } i\in I\setminus\{1\}$$

where μ_1 and μ_2 are two predetermined thresholds.

More complex modeling, including for instance circle arc or rotation component, could also be handled by the same method. This will only lead to other sets of masks to be considered in expression (4).

One advantage of this method is to present no inherent restrictions concerning kinds of edges likely to be successfully handled (in particular, occlusion boundaries) and concerning measurable motion magnitude. The same does not hold for differential methods. The extent of measured motion is directly constrainted by the smoothing extent used to compute the spatial gradient of the image intensity. Therefore, the differential approach is more appropriate for small displacements. Moreover, flow fields are often incorrect near occlusion, since assumptions required for differentiation do not hold any longer in such areas. On the other hand, this maximum-likelihood technique can be CPU-time consuming with a general-purpose computer, but CPU-time can be drastically reduced if an array processor is used.

III COMPLETE ESTIMATION OF THE VELOCITY FIELD ALONG CONTOURS

In the previous section, a procedure has been described which detects moving edges and, at the same time, estimates their local spatial direction and component of their velocity perpendicular to the contour. The goal of the second stage is to compute component of velocity vectors tangent to the contour.

In order to achieve the second stage of the image flow analysis, edge linking is perequisite. To this end, only local spatial directions of detected moving edges are taken into account. One-pixel gaps can be filled up. The linking technique is similar to the one presented in |13|. Then, we get a set of contours, i.e., a set of chains of linked spatio-temporal edges.

To compute the entire velocity field along the contours, the second stage of analysis must combine the local measurements yielded by the first stage. This combination stage is efficient if enough variations in spatial orientation occur along obtained contours. For instance, a straight line contour remains a singular case.

Let $\underline{\omega} = (\omega_x,\omega_y) = \left(\frac{dx}{dt},\frac{dy}{dt}\right)$ be the restriction of $\vec{V} = \left(\frac{dx}{dt},\frac{dy}{dt},1\right)$ to the plane (x,y). Let us consider

$$e_\ell(\underline{\omega}) = \underline{\omega}\ .\ \underline{n}_\ell^T - V_\ell^\perp \tag{6}$$

where $\underline{\omega}$ is the velocity field to be estimated, $\underline{n}_\ell$ is the unitary vector normal to the local edge element at point ℓ, $\underline{n}_\ell=(-\sin\theta_\ell,\cos\theta_\ell)$, $V_\ell^\perp$ is the measured perpendicular component of velocity at point ℓ. $e_\ell(\underline{\omega})$ is supposed to be a stationary random variable.

Then, the measurement of velocity field $\underline{\omega}$ along a given contour C is formulated as the minimization of the following function:

$$J(\underline{\omega}) = \frac{1}{2}\ \mathbb{E}_C(e_\ell(\underline{\omega}))^2 \tag{7}$$

where $\mathbb{E}$ denotes expectation. Motivations for such a criterion can be found in |14|. A stochastic gradient algorithm is used to minimize $J(\underline{\omega})$. The recursive estimation is pursued from one point to its successor in the chain. More precisely, it is expressed as follows:

$$\underline{\omega}_{\ell+1}^T = \underline{\omega}_\ell^T - \gamma\ .\nabla_\omega\ e_\ell(\underline{\omega}_\ell)^T\ .\ e_\ell(\underline{\omega}_\ell) \tag{8}$$

where γ is a gain matrix, and ∇ denotes the gradient.

$$\nabla_\omega \, e_\ell(\underline{\omega}) = (\frac{\partial e_\ell}{\partial \omega_x}, \frac{\partial e_\ell}{\partial \omega_y}), \text{ and}$$

$$\frac{\partial e_\ell}{\partial \omega_x}(\underline{\omega}) = n_\ell^x = -\sin\theta_\ell ;$$

$$\frac{\partial e_\ell}{\partial \omega_y}(\underline{\omega}) = n_\ell^y = \cos\theta_\ell .$$

The initial estimate can be given by $\underline{\omega}_0 = V_0^\perp \underline{n}_0$.

A theoretical proof of such a minimization formulation is shown in |15|. The obtained convergence is quite fast. Two recursion cycles around a given contour C are usually sufficient for a proper estimation of the velocity field $\{\underline{\omega}_\ell, \ell \in C\}$. Moreover, two recursions are performed in parallel, clockwise and counter-clockwise. Then, an average is computed between the two estimates at each point ℓ. Smoothness constraint is not explicitly formulated in the minimization criterion as in |16|, but it is ensured by the recursion along the contour.

IV RESULTS

IV.1 - Results concerning the first analysis stage

Experiments on computer-generated images have been performed in order to warrant the estimation method of moving edges. Different kinds of motion have been considered (translation of the camera along its view axis, object rotation in the image plane). Results are presented in |12|.

The algorithm has also been applied to actual images. Only two successive images are considered for each example reported here. Hence each mask corresponding to coefficients a_j's for each predefined configuration $\Phi_j, j=1,\ldots,G$, divides into two submasks. These G masks are computed once G geometric configurations are chosen. Then they are available when images are processed. Choosing angle ψ_j is equivalent to choosing displacement magnitude $V^\perp_j$ in the direction perpendicular to the local linear edge element whose spatial direction is θ_j. Therefore, the location of the second submask in the second image with respect to location of the first one centered at current point ℓ in the first image is given by $V^\perp_j \, \underline{n}_j$.
Then, the G configurations can also be denoted as $\{((\theta_r, \psi_q), q=1,Q), r=1,R\}$ with $R \times Q = G$. Thus the function CRV can be written as follows:

$$CRV(\ell, \Phi_{rq}) = \left| \sum_{i \in I} CRVi(\ell, \Phi_{rq}) \right|$$

$$= \left| CRV1(\ell, \theta_r) + \sum_{i \in I \setminus \{1\}} CRVi(\ell, \Phi_{rq}) \right| \quad (9)$$

In order to save CPU-time, convolution operations with the whole set of masks, as previously explained, are not actually computed for each point ℓ. If $CRV1(\ell, \theta_r) < a\lambda$, (with e.g. a=0.25), computations corresponding to the evaluation of $CRV(\ell, \Phi_{rq}), q=1,Q$, stop and $CRV(\ell, \Phi_{rq})$, q=1,Q are set to 0.

The first example is extracted from a natural sequence acquired by a camera and depicting an urban scene. Figure 2 shows the first image. The set of masks consists of 66 masks including six possible spatial directions: θ = 0°, 30°, 60°, 90°, 120°, 150° and eleven possible perpendicular displacement magnitudes, $V^\perp = -5,\ldots,-1,0,1,\ldots,5$ (G=66, R=6, Q=11). Submask size is 5x5 pixels. The estimated perpendicular displacement field is presented in figure 3 along with spatial edges, for an image part.

An evaluation of the correctness of the resulting estimation is available. By means of some other tool, motion is known to be three pixels to the left in the whole image except for subparts corresponding to hung linen, some bushes. Quite satisfactory results are obtained.

The second example includes two images of a printer acquired by a CCD camera (Figure 4). Only the printer has been moved from one image to the next, camera and background remain fixed. 84 masks have been considered, that is to say four possible spatial directions θ = 0°, 45°, 90°, 135°, and 21 possible perpendicular displacements $V^\perp = -10,\ldots,-1, 0,1,\ldots,10$. (Of course, metric is adapted when directions other than horizontal and vertical are considered). Determined moving edges are shown in Figure 5 with their perpendicular displacement, which can eventually be none.

IV.2 - Results concerning the second analysis stage

Two sets of experiments are presented involving computer-generated images including a single polygonal object and two kinds of motion: uniform translation and in-plane rotation. Superimposed silhouettes of the object in two successive positions are shown for both cases, respectively in Fig. 6a and 7a. Of course, the method is not restricted to such cases.
In Fig. 6b and 7b is drawn the perpendicular displacement field. It has been estimated using the algorithm presented in this paper.
Fig. 6c and 7c show the resulting complete displacement field after two recursive estimation cycles around the boundary. The last example points out that varying displacement field can be successfully handled.

V FUTURE WORK

Future research directions mainly include:
- corner displacement estimation (as complementary processing after linking)
- detection of possible motion boundaries along contours in parallel with the recursive estimation (e.g., this may happen if a boundary portion is an occlusion one) in order to reinitialize the recursion.

REFERENCES

|1| T.S.Huang (Editor), "Image Sequence Processing and Dynamic Scene Analysis", Proc. of NATO Advanced Study Institute at Braunlage, W-Germ., Springer-Verlag, 1983.

|2| R.M.Endlich, D.E.Wolf, D.J.Hall, and A.E.Brain, "Use of a Pattern Recognition Technique for Determining Cloud Motions from Sequences of satellite photographs", Jal of Applied Meteorology, Vol.10, Feb. 1971, pp 105-117.

|3| J.R.Jain, and A.K.Jain, "Displacement Measurement and Its Application in Interframe Image Coding", IEEE Trans. on Communications, Vol. COM-29, n°12, Dec. 1981, pp 1799-1808.

|4| R.Paquin, and E.Dubois, "A Spatio-Temporal Gradient Method for Estimating the Displacement Field in Time-Varying Imagery", Computer Vision, Graphics, and Image Proc., 21, 1983, pp 205-221.

|5| J.K.Tsotsos, J.Mylopoulos, H.D.Covvey, and S.W. Zucker, "A Framework for Visual Motion Understanding", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.PAMI-2, n°6, Nov. 1980, pp 563-573.

|6| P.Rives, and L.Marcé, "Use of Moving Vision Sensors to Control Robots in an Unknown Universe", Proc. of ROVISEC 5, Amsterdam, Oct. 1985, pp 165-176.

|7| A.Mitiche, "On Kineopsis and Computation of Structure and Motion", 2d IEEE Comp. Vision Workshop on Rep. and Control, Annapolis, May 1984.

|8| G.Adiv, "Determining Three-Dimensional Motion and Structure from Optical Flow Generated by Several Moving Objects", IEEE Trans. on PAMI, Vol.7, n°4, July 1985, pp 384-401.

|9| S.T.Barnard, and W.B.Thompson, "Disparity Analysis of Images", IEEE Trans. on PAMI, Vol.2, n°4, 1980, pp 333-340.

|10| B.G.Schunck, and B.K.P.Horn, "Constraints on Optical Flow Computation", Proc. of Pattern Recognition and Image Proc. Conf., Dallas, Aug. 1981, pp 205-210;

|11| H.H.Nagel, "Constraints for the Estimation of Displacement Vector Fields from Image Sequences", Proc. of IJCAI, 1983, pp 945-951.

|12| P.Bouthemy, "Estimation of Edge Motion Based on Local Modeling", Proc. of SPIE Conf. on Computer Vision for Robots, Cannes, Dec. 1985.

|13| R.Nevatia, and K.R.Babu, "Linear Feature Extraction and Description", Computer Graphics and Image Proc., 13, 1980, pp 257-269.

|14| P.Bouthemy, "Un Nouveau Schéma d'Estimation du Champ des Vitesses sur les Contours dans une Séquence d'Images", Proc. of CESTA Second Image Symposium, Nice, Avril 1986.

|15| A.Benveniste, M.Goursat, and G.Ruget, "Analysis of Stochastic Approximation Schemes with Discontinuous and Dependent Forcing Terms with Application to Data Communication Algorithms", IEEE Trans. on Automatic Control, Vol.AC-25, n°6, Dec. 1980, pp 1042-1058.

|16| E.C.Hildreth, "Computations Underlying the Measurement of Visual Motion", Artif. Intell., 23, 1984, pp 309-354.

ACKNOWLEDGEMENT

The author would like to thank Albert Benveniste for useful comments.

Fig.2: *First image of the "house" sequence*

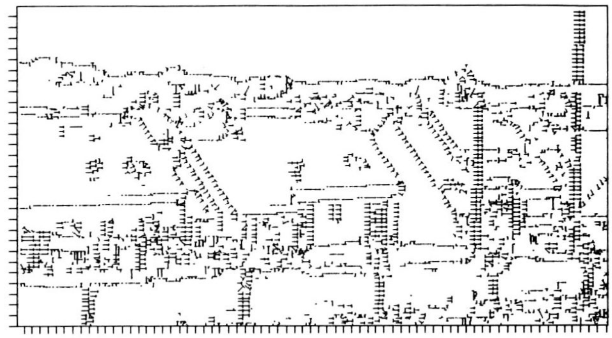

Fig.3: *Determined spatio-temporal edges with their perpendicular displacement (axes are sampled each 5 pels)* $\lambda=2500$, $\mu_1=0.8$, $\mu_2=1.2$

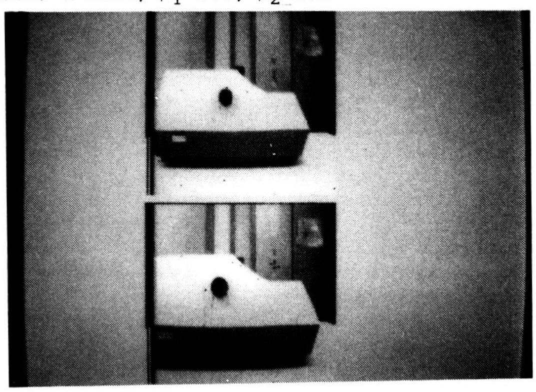

Fig.4: *"Printer" sequence*

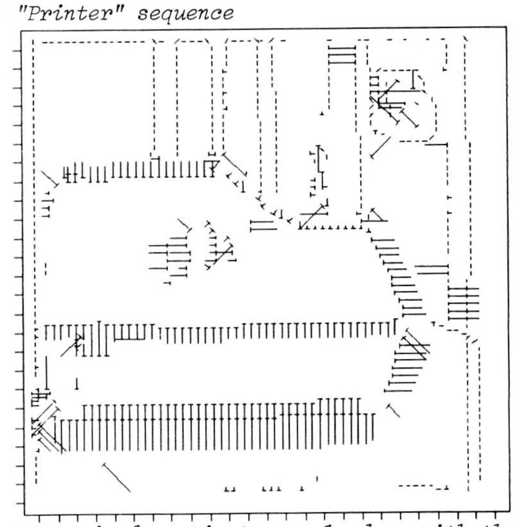

Fig.5: *Determined spatio-temporal edges with their perpendicular displacement* $\lambda=3000$, $\mu_1=0.75$, $\mu_2=1.25$

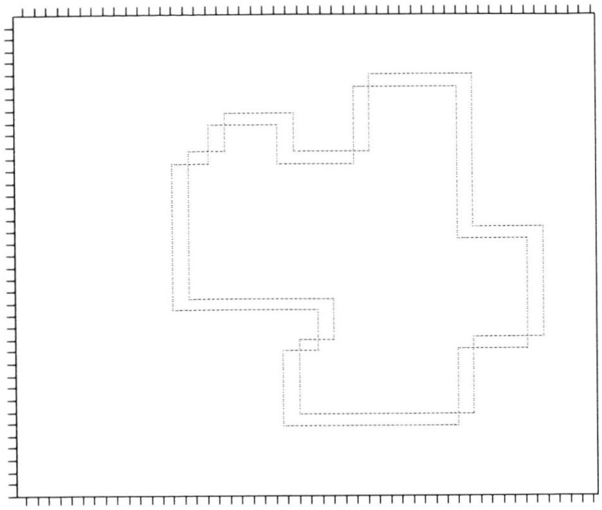

Fig.6a: *Superimposed silhouettes of polygonal object 1*

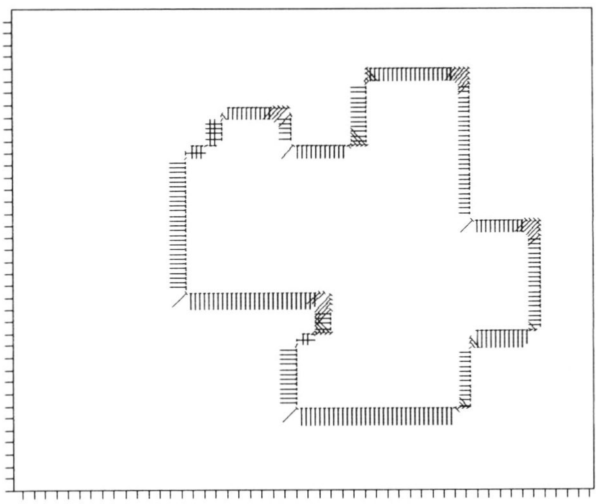

Fig.6b: *Perpendicular displacement field*

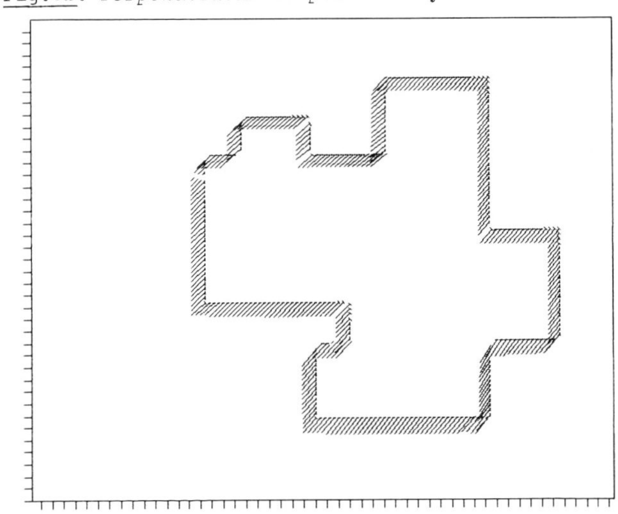

Fig.6c: *Resulting complete displacement field*

$$\gamma = \begin{pmatrix} 0.03 & 0.01 \\ 0.01 & 0.03 \end{pmatrix}$$

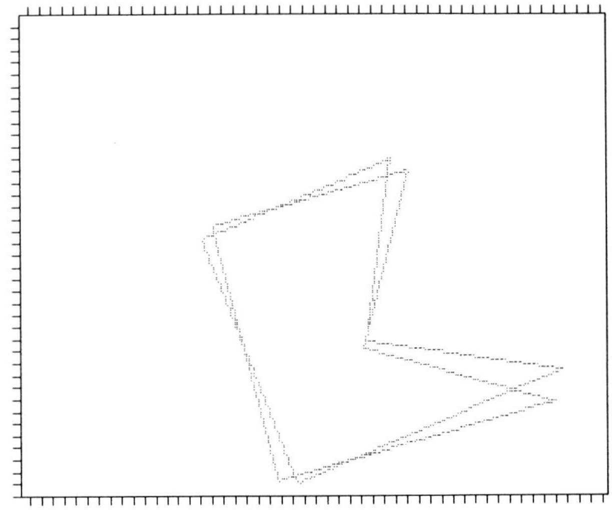

Fig.7a: *Superimposed silhouettes of polygonal object 2*

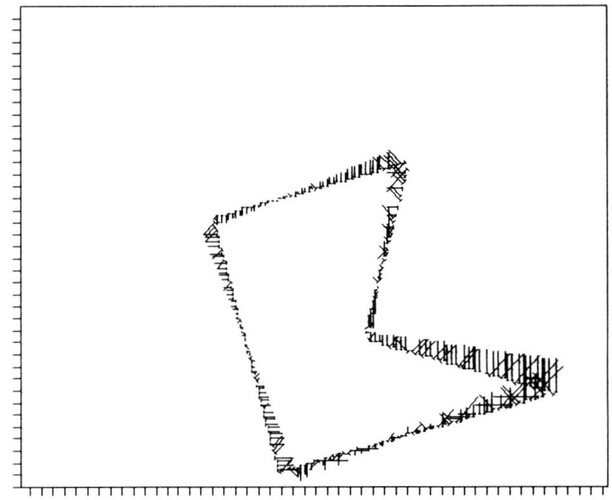

Fig.7b: *Perpendicular displacement field*

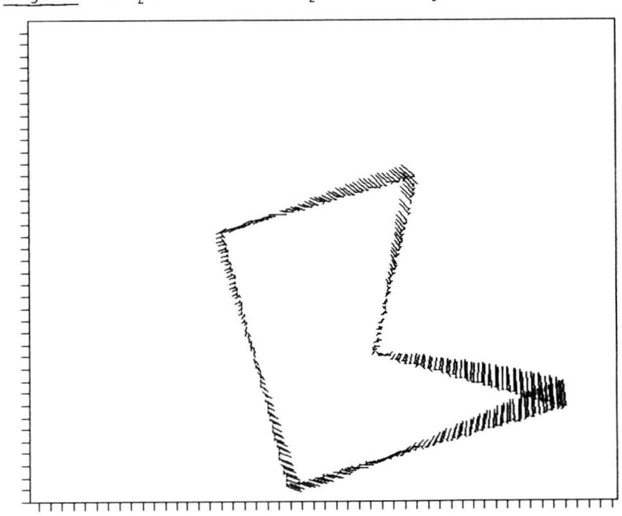

Fig.7c: *Resulting complete displacement field*

$$\gamma = \begin{pmatrix} 0.04 & 0.01 \\ 0.01 & 0.04 \end{pmatrix}$$